5-1993

Molecular basis of the Ribulose-1,5-Bisphosphate Carboxylase/ Oxygenase Activase Mutation in Arabidopsis Thaliana is a Guanine-To-Adenine Transition at the 5'-Splice Junction of Intron 3

Beverly M. Orozco
University of Illinois at Urbana-Champaign

C Robertson McClung
Dartmouth College

Jeffrey M. Werneke
University of Illinois at Urbana-Champaign

William L. Ogren
University of Illinois at Urbana-Champaign

Plant Physiol. (1993) 102: 227–232

Molecular Basis of the Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Activase Mutation in *Arabidopsis thaliana* Is a Guanine-to-Adenine Transition at the 5′-Splice Junction of Intron 3[1]

Beverly M. Orozco, C. Robertson McClung, Jeffrey M. Werneke[2], and William L. Ogren*

Department of Plant Biology, University of Illinois at Urbana-Champaign (B.M.O., J.M.W., W.L.O); Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire (C.R.M.); and Photosynthesis Research Unit, Agricultural Research Service, United States Department of Agriculture, Urbana, Illinois 61801-3838 (W.L.O.)

Analysis of the ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) activase gene and gene products from *Arabidopsis thaliana* wild-type plants and the Rubisco activase-deficient mutant strain showed that the *rca* mutation caused GT to be changed to AT at the 5′-splice junction of intron 3 in the six-intron pre-mRNA. Northern blot analysis, genomic and cDNA sequencing, and primer extension analysis indicated that the mutation causes inefficient and incomplete splicing of the pre-mRNA, resulting in the accumulation of three aberrant mRNAs. One mutant mRNA was identical with wild-type mRNA except that it included intron 3, a second mRNA comprised intron 3 and exons 4 through 7, and the third mRNA contained exons 1 through 3. The G-to-A transition is consistent with the known mechanism of mutagenesis by ethyl methanesulfonate, the mutagen used to create the Rubisco activase-deficient strain.

Rubisco initiates the photosynthetic carbon reduction and photorespiratory carbon oxidation pathways by catalyzing the carboxylation or oxygenation of RuBP (Ogren, 1984). Rubisco must be converted to an active state for catalytic competence by the addition of CO_2 and Mg^{2+} at a Lys residue near the active site (Lorimer and Miziorko, 1980). In vivo, Rubisco activation state is regulated by Rubisco activase, a soluble chloroplast enzyme (Salvucci et al., 1985; Portis et al., 1987). Rubisco activase was identified by characterizing a high-CO_2-requiring mutant in *Arabidopsis*, designated *rca*, that was unable to activate Rubisco in the light (Somerville et al., 1982). This mutant was recovered by treating *Arabidopsis* seeds with the chemical mutagen EMS and then screening the M_2 population for mutants that survived at elevated levels of CO_2 but senesced when grown in air. Rubisco deactivated when *rca* mutant plants were placed in air, but the purified enzyme appeared to be identical with wild type, indicating that the mutation did not affect Rubisco itself. Comparison of the soluble proteins from mutant and wild-type plants by two-dimensional PAGE revealed that two polypeptides were missing in the mutant (Salvucci et al., 1985). The polypeptides were subsequently purified and shown to promote the activation of Rubisco at physiological concentrations of CO_2, Mg^{2+}, and RuBP (Portis et al., 1987).

Rubisco activase was found by immunological screening to be present in a range of plant species (Salvucci et al., 1987). Most species examined contain two Rubisco activase polypeptides with approximate sizes of 41 and 45 kD. Werneke et al. (1989) showed that the two polypeptides are encoded by a single nuclear gene in spinach (*Spinacea oleracea* L.) and *Arabidopsis*. Alternative splicing of the spinach Rubisco activase transcript leads to the production of two mRNA populations that differ by a short auxiliary exon (auxon) present in the larger message. The auxon contains a stop codon, which results in early termination of translation and synthesis of the shorter polypeptide. Alternative splicing of the transcript to remove the auxon allowed translation of an additional 37 amino acids at the C-terminal end, and the 45-kD polypeptide is synthesized. In *Arabidopsis*, alternative splicing of the Rubisco activase transcript leads to the synthesis of 42- and 46-kD polypeptides. In barley (*Hordeum vulgaris*) there are two Rubisco activase genes, which occur in tandem, one gene encoding two mRNAs by alternative splicing and the second gene encoding a third mRNA (Rundle and Zielinski, 1991).

Rubisco activase polypeptides were not found and did not appear to be synthesized in the *rca* mutant (Salvucci et al., 1985; Werneke et al., 1988). However, when northern blot analysis was used with a 0.5-kb Rubisco activase cDNA probe to the 3′ end of the mRNA, two bands of approximately 1.8 and 2.0 kb in length were revealed with mutant *Arabidopsis* mRNA, whereas wild-type *Arabidopsis* mRNA contained a single 1.9-kb band (Werneke et al., 1988). All Rubisco activase cDNA clones recovered from an *Arabidopsis rca* mutant library contained normal 3′ sequences, but no full-length

[1] C.R.M. was supported by National Science Foundation grant DMB-9005345 and American Cancer Society grant JFRA-253.

[2] Present address: Abbott Laboratories, Diagnostics Division, One Abbott Park Road, Abbott Park, IL 60064-3500.

* Corresponding author; fax 1–217–244–4419.

Abbreviations: EMS, ethyl methanesulfonate; PCR, polymerase chain reaction; *rca*, mutant rubisco activase gene; *Rca*, wild-type rubisco activase gene; RuBP, ribulose-1,5-bisphosphate.

cDNAs were isolated. These results suggested that a mutation at the 5′ end of the gene disrupted normal pre-mRNA splicing, which in turn prevented synthesis of the Rubisco activase polypeptides (Werneke et al., 1988). To identify the lesion creating the rca phenotype, genomic and cDNA clones were isolated from the mutant and the sequences were compared with wild type. In addition, northern analysis was performed to characterize further the mRNA products resulting from the mutation.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Arabidopsis thaliana (L.) Heynh. Columbia wild-type and *rca* mutant plants (Somerville et al., 1982) were grown in pots half filled with perlite and covered with Jiffy Mix[3] (a vermiculite, peat moss, soil mixture). Seeds were evenly distributed on the soil surface using a 0.1% agarose suspension in a pasteur pipette, and the pots were covered with plastic wrap. When the first true leaves expanded the plastic wrap was gradually removed over several days. Plants were grown at 22°C and supplemented with 100% CO_2 to maintain an atmosphere of about 1% CO_2.

Construction of the *rca* Genomic DNA Library

Nuclei were isolated from the *rca* mutant *Arabidopsis* by grinding leaves in an extraction buffer (250 mM Suc, 200 mM NaCl, 10 mM $MgCl_2$, 5 mM DTT, and 100 mM Tris [pH 9.0]), filtering through Miracloth, and centrifuging at 9000*g* for 10 min. The pellet was resuspended in 50 mL of buffer containing 100 mM EDTA (pH 8.0), 0.5% sarcosyl, and 200 μg/mL of proteinase K and incubated at 50°C for 1 h.

The genomic DNA was partially digested with *Sau*3A for 1.5 to 2.5 min, extracted with phenol, and precipitated with 0.2 M NaCl in 2.0 volumes of ethanol. The digested DNA was cloned into the *Bam*HI site of EMBL4 and packaged in vitro using the Gigapack Plus Kit (Stratagene). The library contained approximately 30,000 clones. DNA probes were prepared by random hexamer labeling using [α-^{32}P]dATP. Approximately 25 to 50 ng of DNA were denatured by boiling and then chilled on ice. The labeled strands were synthesized by adding 0.05 A_{260} units of random hexamers (Pharmacia), 50 μCi of 6000 Ci mmol^{-1} of [α-^{32}P]dATP, and 0.5 units of Sequenase (United States Biochemical) in a reaction buffer containing 100 mM Mops-NaOH (pH 6.6), 5 mM $MgCl_2$, 1 mM DTT, 50 μg mL^{-1} of BSA, and 20 μM each dGTP, dCTP, and dTTP. The reaction incubated at 37°C for 1 h. Unincorporated nucleotides were then removed by G-100 Sephadex column chromatography.

Hybridization Screening of Libraries

Genomic and cDNA libraries were screened by hybridization essentially as described by Maniatis et al. (1982). The *rca* mutant libraries were plated with top agar using host *Escherichia coli* strain Y1088 and incubated at 37°C overnight. Phage DNA was bound to nitrocellulose and denatured. Filters were prehybridized for 30 to 60 min at 65°C in 675 mM NaCl, 5 mM EDTA, and 50 mM NaH_2PO_4/Na_2HPO_4 (pH 7.4), 5× Denhardt's solution (0.1% Ficoll, 0.1% PVP, and 0.1% BSA), 0.5% SDS, and 200 μg mL^{-1} of denatured salmon sperm DNA. Approximately 10^6 cpm of probe per mL of hybridization solution was denatured in a boiling water bath and added to the same solution. The filters were hybridized at 65°C for 10 to 24 h.

Isolation of *rca* cDNA Clones

Total RNA was isolated from a single *rca* mutant *Arabidopsis* plant as described by Verwoerd et al. (1989). Total RNA was then hybridized with the synthetic oligonucleotide 5′-TCAGCTGCCTCACGGTACCTCTGACGGATAAGCTT-3′ at 48°C overnight in a buffer containing 400 mM NaCl, 40 mM Pipes (pH 6.8), 1 mM EDTA, 50% formamide, and 30 units of RNA-guard (Promega Biotec). cDNA was synthesized from the hybrids in primer extension buffer (50 mM Tris HCl [pH 8.0], 70 mM KCl, 6 mM $MgCl_2$, 10 mM DTT, 500 μM deoxynucleotide triphosphate, 250 μg mL^{-1} of BSA, 30 units of RNA-guard), and 5 units of avian myeloblastosis virus reverse transcriptase (BRL) for 1 h at 37°C (Werneke et al., 1989). The reaction was stopped with 20 mM EDTA and the cDNA precipitated with 0.3 M sodium acetate and 2 volumes of ethanol. The pellet was resuspended in water, and 0.1 volume was used for PCR amplification (Kawasaki, 1990) of the *rca* cDNA using the GeneAmp Kit (Perkin-Elmer Cetus). The cDNA was amplified for 30 cycles with 50 pmol of each primer (5′ primer 5′-GGATCCGACGCATATCTCGTCG-CAGTCTTGAGATATGGCC-3′ and 3′ primer listed above) under the following conditions: denaturation for 2 min at 94°C, annealing for 2 min at 60°C, and extension for 4 min at 72°C. The PCR product was cloned into pTZ-18U *Bam*HI and *Kpn*I restriction sites.

Northern Blot Analysis

Total RNA was fractionated on 1.2% agarose gels containing formaldehyde and transferred to nitrocellulose in 3 M NaCl 3 M sodium citrate (pH 7.0) by capillary blot overnight (Maniatis et al., 1982). The filters were dried and baked for 1 h at 90°C. Prehybridization and hybridization procedures were as described above for screening libraries. Four probes were prepared: a full-length cDNA (pARAB18, Werneke and Ogren, 1989); a 5′ fragment of cDNA containing exons 1 through 3 (nucleotides 1343–1449 and 1925–2241 in Fig. 1); a 3′ fragment of cDNA that included exons 6 and 7 (nucleotides 3370–3570 and 3681–3890 in Fig. 1); and a cDNA fragment containing exon 3 (nucleotides 2241–2249 and 2342–2429 in Fig. 1).

Primer Extension Analysis

The synthetic oligonucleotide 5′-GGATTTACCTTGAC-CCGCGCCTCCCCAAAC-3′ was radiolabeled with [γ-^{32}P]-ATP using the TaqTrack Sequencing Kit (Promega). Reverse

[3] Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture and does not imply its approval to the exclusion of other products or vendors that may also be suitable.

```
   1  CCAACATTAT TAATTTATAA AGGTTTTACT GTACTTCATT TTCTAACCTC AGTATGTTAA AAGTCATGTA TGTGTACTAA GAATCAAACC TGGCACGCAA
 101  ACTTTTGCTA ATTTTGTCAC ATACAGTTAA CATTTAGCAT TGTGCGAGTA GCTCCACCAC ATTATTATCA ATCGTTTGGT TCTTCAACAT GTCTTATATA
 201  TGGATCAATA ATAGTTTAAT ATCATATCAG TAATCTTTGT TCTCTTGTCT TGCATGTAGA ACATGTACAT ATTGCTTCGT CGGCAGCGAA TGAAGCCTAA
 301  AATTAAAATT TGGTTTTGTC CATTGATGAT ATGATGGCTC GTTCCACATG AGAGAATGGT TGAGAAAACT GAAAATAATG AAGAGGATGA GAAATTATAA
 401  TGATCAGACA AGTGAAGTGT TTCATTTTAT AGCTTAGTGT AGAAAACGAT TGCTTGTGAT TGTGAATATA TAGTTTTAGA TGAGTTGTTT TTATAGTAAC
 501  TTTGAGGGAA CAATAAACTA TGATATAGTT ATCATTTTTA TTTTAAAACA AGTAAATTCA TATGAATTTC TGATATATTC ATATTTTCAC TATGAATTTA
 601  AACACTATAA TCTTATTTCA TATTTTAGCA TATGATATTT TGCATGATGC TATCTTAGCC TCGACTATCT TTACTTCATC TTCTAAACTA TATATTTATC
 701  AAATCTTACG TATGCATTGC ATCACTTGTT TTTACATTTT GTGTAAAGAA AAAGTTGCAT TAATTAGTTT TTAGCGGTTT AGTACAATAT GGTTAATTTG
 801  ATAACCACAT ACTTGATACC TGTTGTATCT AGAGTTGTAT ATAACAAAAT TAGAATTATA TTCTATTGGC TAAGCCAATT TCTCTCCACC AATCAGTTTA
 901  CATCGGTTAT TTGGAACATT ATTATCGATA AGCAATGTCA TTATCCAAAA TTTTGGAATG TTATCGTATG AAGAAGAGTG TAGAAGTATC TTGATCACAT
1001  ATGAATGGCT TACCTCAATC CTCAAAAAAA AAAAAAACA ATGATTTTCC AAAGCATATG TGGTAATAAA TAATGTTATA ATCTTCACTT GAAAAAAAAA
1101  ACAAATCCAT TGGCACAGAT TAGCCATTTG ACGACAACCT CATAAACATA TTCCACGTGG ACCAATCAAA ATAGAATCCT CCTAATCATC TTCCTCGTGG
1201  CAACTTGGCC AATCCGCATC GTGTGGCGAT CAGAGAGTTA AGCCTTGAAG ACGAAGAGAT AACGAATTGG TTGTTGATCA CTCGCTTTAT AAATCTCTCA
1301  GTTTCTTGCT CACACCAACA TCTCTCTAAG CTTCTTCTTC TACCAATCTA ATTCCTCTCT TCAGCTTCTT GTGTTGTGAC GCATACTCGT CGCAGTCTTG
1401  AGAT ATG GCC GCC GCA GTT TCC ACC GTC GGT GCC ATC AAC AGA GCT CCG  GTACACTTCA ATCTCCATTA GTGGTTTAAT GTTACTTTTA
          Met Ala Ala Ala Val Ser Thr Val Gly Ala Ile Asn Arg Ala Pro                  INTRON I
1490  ACATAATCTT AACGCTTTGT GCTTATCCAC CACTTGAATA GCTTTGGTTT TTACAAATAT AGCTCAAGAA TTTGATCGTT TGAATCACTT TGAGATTCTC
1590  TGACAGAACA GAAGAATCAG GACTTTAACT AGGAAATCTA AAGATACTTT GATGCTTCTT TTATTCTCTA CTTGTGAACT ATAGATTTAG TAGATGTGTT
1690  TAAATCGATG TATAGTAATT GAACTGAAAC TGTAGAGACT ACTACTACTC TTTAGTCTTC AGGACTAGTC TATACCTTGA GCTATGTGAA CCAAATTAAA
1790  GTTTATTCCC CGGTTTGGTT CGGTTCAGTT CCTGATTGAG CGCCTTTTAT CCTTAGGAAT AGTCTCTAAA CCGGTTCTTG AACCATATAA ACCAAACTTG
1890  AATTCATGAT TGGTGATTGA ACTGTTTTTG TGTGTGTGTT GCAG    TTG AGC TTG AAC GGG TCA GGA TCA GGA GCT GTA TCA GCC CCA GCT
                                                        Leu Ser Leu Asn Gly Ser Gly Ser Gly Ala Val Ser Ala Pro Ala
1979  TCA ACC TTC TTG GGA AAG AAA GTT GTA ACT GTG TCG AGA TTC GCA CAG AGC AAC AAG AAG AGC AAC GGA TCA TTC AAG GTG TTG
      Ser Thr Phe Leu Gly Lys Lys Val Val Thr Val Ser Arg Phe Ala Gln Ser Asn Lys Lys Ser Asn Gly Ser Phe Lys Val Leu
2063  GCT GTG AAA GAA GAC AAA CAA ACC GAT GGA GAC AGA TGG AGA GGT CTT GCC TAC GAC ACT TCT GAT GAT CAA CAA GAC ATC ACC
      Ala Val Lys Glu Asp Lys Gln Thr Asp Gly Asp Arg Trp Arg Gly Leu Ala Tyr Asp Thr Ser Asp Asp Gln Gln Asp Ile Thr
2147  AGA GGC AAG GGT ATG GTT GAC TCT GTC TTC CAA GCT CCT ATG GGA ACC GGA ACT CAC CAC GCT GTC CTT AGC TCA TAC GAA TAC
      Arg Gly Lys Gly Met Val Asp Ser Val Phe Gln Ala Pro Met Gly Thr Gly Thr His His Ala Val Leu Ser Ser Tyr Glu Tyr
2231  GTT AGC CAA GGC CTT AGG CA         GTAAGTTCTC ATCTCTCTAG ATCTCTCATT AGTTCCATAA AGATTTTGGA GTCTTTGGTT TTAAAAATTG
      Val Ser Gln Gly Leu Arg Gl                                    INTRON II
2321  TCTTTACTTC TTGAAATATAG    G TAC AAC TTG GAC AAC ATG ATG GAT GGG TTT TAC ATT GCT CCT GCT TTC ATG GAC AAG CTT GTT
                               n Tyr Asn Leu Asp Asn Met Met Asp Gly Phe Tyr Ile Ala Pro Ala Phe Met Asp Lys Leu Val
2406  GTT CAC ATC ACC AAG AAC TTC TTG ACT CTG CCT AAC ATC AAG  GTATAATTTA AAGTTTTTCT TTTGAGTATT TTTTGGTGCA ATCTTGACTT
      Val His Ile Thr Lys Asn Phe Leu Thr Leu Pro Asn Ile Lys                       INTRON III
2498  ATTACACAAC TTTATTGATC TTTAAATGCC TGTTTTTTTT GTCTCATAG  GTT CCA CTT ATT TTG GGT ATA TGG GGA GGC AAA GGT CAA GGT
                        ↑                                   Val Pro Leu Ile Leu Gly Ile Trp Gly Gly Lys Gly Gln Gly
2589  AAA TCC TTC CAG TGT GAG CTT GTC ATG GCC AAG ATG GGT ATC AA         GTGAGTCCAT TTCCCAAATC TCCGATCTTG ACCCGGTTTC
      Lys Ser Phe Gln Cys Glu Leu Val Met Ala Lys Met Gly Ile As                        INTRON IV
2673  TGTTTAGCTC TCTTTCGACA AGAATCTAAA AAAGTTTGAA TCTTTTTGCAG   C CCA ATC ATG ATG AGT GCT GGA GAG CTT GAG AGT GGA AAC
                                                              n Pro Ile Met Met Ser Ala Gly Glu Leu Glu Ser Gly Asn
2764  GCA GGA GAA CCC GCA AAG CTT ATC CGT CAG AGG TAC CGT GAG GCA GCT GAC TTG ATC AAG AAG GGA AAG ATG TGT TGT CTC TTC
      Ala Gly Glu Pro Ala Lys Leu Ile Arg Gln Arg Tyr Arg Glu Ala Ala Asp Leu Ile Lys Lys Gly Lys Met Cys Cys Leu Phe
2848  ATC AAC GAT CTT GAC GCT GGT GCG GGT CGT ATG GGT GGT ACT ACT CAG TAC ACT GTC AAC AAC CAG ATG GTT AAC GCA ACA CTC
      Ile Asn Asp Leu Asp Ala Gly Ala Gly Arg Met Gly Gly Thr Thr Gln Tyr Thr Val Asn Asn Gln Met Val Asn Ala Thr Leu
2932  ATG AAC ATT GCT GAT AAC CCA ACC AAC GTC CAG CTC CCA GGA ATG TAC AAC AAG GAA GAG AAC GCA CGT GTC CCC ATC ATT TGC
      Met Asn Ile Ala Asp Asn Pro Thr Asn Val Gln Leu Pro Gly Met Tyr Asn Lys Glu Glu Asn Ala Arg Val Pro Ile Ile Cys
3016  ACT GGT AAC GAT TTC TCC ACC CTA TAC GCT CCT CTC ATC CGT GAT GGA CGT ATG GAG AAG TTC TAC TGG GCC CCG ACC CGT GAA
      Thr Gly Asn Asp Phe Ser Thr Leu Tyr Ala Pro Leu Ile Arg Asp Gly Arg Met Glu Lys Phe Tyr Trp Ala Pro Thr Arg Glu
3100  GAC CGT ATC GGT GTC TGC AAG GGT ATC TTC AGA ACT GAC AAG ATC AAG GAC GAA GAC ATT GTC ACA CTT GTT GAT CAG TTC CCT
      Asp Arg Ile Gly Val Cys Lys Gly Ile Phe Arg Thr Asp Lys Ile Lys Asp Glu Asp Ile Val Thr Leu Val Asp Gln Phe Pro
3184  GGT CAA TCT ATC G  GTAAGAAAGA TTCTTACTAC CCTGAAAATG TTCTATATAG GACCTCGTGC AGAATCTGAT ATGTGTTTTT GGTGTGGACAG  AT
      Gly Gln Ser Ile A                                             INTRON V                                      sp
3280  TTC TTC GGT GCT TTG AGG GCG AGA GTG TAC GAT GAT GAA GTG AGG AAG TTC GTT GAG AGC CTT GGA GTT GAG AAG ATC GGA AAG
      Phe Phe Gly Ala Leu Arg Ala Arg Val Tyr Asp Asp Glu Val Arg Lys Phe Val Glu Ser Leu Gly Val Glu Lys Ile Gly Lys
3364  AGG CTC GTT AAC TCA AGG GAA GGA CCT CCC GTG TTC GAG CAA CCC GAG ATG ACT TAT GAG AAG CTT ATG GAA TAC GGA AAC ATG
      Arg Leu Val Asn Ser Arg Glu Gly Pro Pro Val Phe Glu Gln Pro Glu Met Thr Tyr Glu Lys Leu Met Glu Tyr Gly Asn Met
3448  CTT GTG ATG GAA CAA GAG AAT GTC AAG AGA GTC CAA CTT GCC GAG ACC TAC CTC AGC CAG GCT GCT TTG GGA GAC GCA AAC GCT
      Leu Val Met Glu Gln Glu Asn Val Lys Arg Val Gln Leu Ala Glu Thr Tyr Leu Ser Gln Ala Ala Leu Gly Asp Ala Asn Ala
a3532  GAC GCC ATC GGC CGC GGA ACT TTC TAC G         GTAAAACAGA GGTTTGAAAC TCGTTTGTTT TTTGGTTTAG TTGGCCACTG GGTATGAAAA
       Asp Ala Ile Gly Arg Gly Thr Phe Tyr G                                 INTRON VIa
b3532  GAC GCC ATC GGC CGC GGA ACT TTC TAC GGT AAA ACA GAG   GTTTGAAAC TCGTTTGTTT TTTGGTTTAG TTGGCCACTG GGTATGAAAATGTA
       Asp Ala Ile Gly Arg Gly Thr Phe Tyr Gly Lys Thr Glu                          INTRON VIb
a3620  AAAGAG AAAGAGAAGA AAATGGAATT AGGTTTTCTG AAGTTAAACT TACTTGACTT GCAG   GA AAA GGA GCC CAG CAA GTA AAC CTG CCA GTT
                                                                             ly Lys Gly Ala Gln Gln Val Asn Leu Pro Val
b3620  AAAGAG AAAGAGAAGA AAATGGAATT AGGTTTTCTG AAGTTAAACT TACTTGACTT GCAG   GAA AAG GAG CCC AGC AAG TAA    ACCTGCCAGTT
                                                                            Glu Lys Glu Pro Ser Lys Stop
a3716  CCT GAA GGG TGT ACT GAT CCT GTG GCT GAA AAC TTT GAT CCA ACG GCT AGA AGT GAC GAT GGA ACC TGT GTC TAC AAC TTT TGA
       Pro Glu Gly Cys Thr Asp Pro Val Ala Glu Asn Phe Asp Pro Thr Ala Arg Ser Asp Asp Gly Thr Cys Val Tyr Asn Phe Stop
b3716  CCTGAAGGGT GTACTGATCC TGTGGCTGAA AACTTTGATC CAACGGCTAG AAGTGACGAT GGAACCTGTG TCTACAACTT TTGA
 3800  GCAATATTAT CCTGCTTATT AATTTGCTGT TTTACTCCTA TTGTCTCTTT TGGTTTATTT TTCTCCTTTG TGTAATTGTG GATTGGATCT TGTCCTCTTT
 3900  TGTCCTCTTT TGTTCCCTTT TTTTTTTTTT ATGATGTACA ACACATTGGT AATTTAAAAT TGCCTTGTCA TAACCGAATT TGGCC
```

Figure 1. Genomic sequence of the *A. thaliana* Rubisco activase gene. Coding regions for the two polypeptides generated by alternative splicing of intron 6 are designated a (42 kD) and b (46 kD). The putative branch site for lariat formation in intron 3 is underlined (nucleotides 2512–2516), and the primer extension termination site is indicated by an arrow (nucleotide 2518).

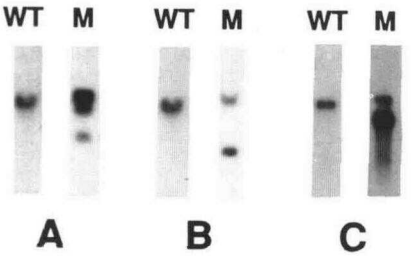

Figure 2. Northern blot analysis of Rubisco activase transcripts in wild-type (WT) and *rca* mutant (M) mRNA. A, Full-length probe; B, 5′ probe; C, 3′ probe. The nucleotide composition of the probes is given in "Materials and Methods."

transcription and gel fractionation were done as described by Werneke (1989).

RESULTS

Genomic Sequence of the Rubisco Activase Gene

The nucleotide sequence for a 3974-nucleotide region of the *Arabidopsis* genome containing the *Rca* gene is given in Figure 1. By comparison with the cDNA sequence (Werneke and Ogren, 1989), the gene contains six introns, as do the spinach gene (Werneke et al., 1989) and one of the two barley genes (Rundle and Zielinski, 1991). As reported previously (Werneke et al., 1989), alternative splicing can occur in intron 6 to give the 42- and 46-kD polypeptides seen in extracts of wild-type *Arabidopsis* leaves (Werneke et al., 1988).

Northern Blot Analysis

Arabidopsis wild-type and *rca* mutant mRNA were subjected to northern blot analysis with three different probes. A full-length cDNA probe revealed three aberrant transcripts in the mutant, approximately 0.5, 1.8, and 2.0 kb in length, and a single 1.9-kb transcript in the wild-type (Fig. 2A). A 5′-specific probe containing exons 1 and 2 annealed with the 0.5- and 2.0-kb transcripts (Fig. 2B), whereas a 3′-specific probe containing exons 6 and 7 used previously (Werneke et al., 1988) annealed with the 1.8- and 2.0-kb transcripts (Fig. 2C). Additionally, a probe specific for exon 3 also hybridized with the 0.5-kb mRNA (data not shown). These results indicate that the lesion in the *rca* mutant disrupts normal processing of the Rubisco activase pre-mRNA. Two aberrant transcripts were produced that contained either the 5′ end (0.5 kb) or 3′ end (1.8 kb) of the wild-type mRNA but not both ends. The presence of the 2.0-kb transcript, which is larger than wild-type mRNA, indicates that an incompletely spliced precursor also accumulates.

Analysis of cDNA Clones

Four cDNA clones positive for Rubisco activase were isolated from the *rca* mutant library. These were subcloned into pTZ-18U and sequenced. Each of the cDNAs contained wild-type sequences from the 3′ end of the message and intron 3 sequences up to 13 nucleotides from the 5′ junction of the intron. No cDNAs containing regions upstream from intron 3 were isolated.

Reverse transcription and PCR amplification were used to isolate cDNAs from the large, 2.0-kb transcript observed in the northern analysis. A PCR product corresponding to nucleotides 45 to 697 in the wild-type cDNA pARAB18 (Werneke and Ogren, 1989; nucleotides 1379–1404, exons 1–3, and nucleotides 2714–2797 in Fig. 1) was generated with total RNA from a single *rca* mutant *Arabidopsis* plant. The fragment was cloned into *Hin*dIII and *Kpn*I sites of pTZ-18U and sequenced. The cDNA was identical with wild type in this region except that intron 3 was retained and a G-to-A transition was observed in the conserved GT dinucleotide at the 5′-splice junction (data not shown). Introns 1 and 2 were not present, indicating that the amplification product was derived from mRNA and not from genomic DNA.

Analysis of Genomic Clones

Three positive Rubisco activase clones were isolated from the *rca* mutant *Arabidopsis* genomic library. The inserts were subcloned into pTZ-18U and partially mapped by restriction analysis. One 6.5-kb clone produced a restriction fragment pattern identical with that of the wild-type genomic clone. *Eco*RI and *Hin*dIII digestion fragments that hybridized with the wild-type cDNA were then subcloned and sequenced. As observed with the cDNAs, the genomic clone contained a G-to-A transition at the 5′-splice junction of intron 3 (Fig. 3).

Primer Extension Analysis

Reverse transcription was performed with total RNA from the *rca* mutant by using a synthetic oligonucleotide down-

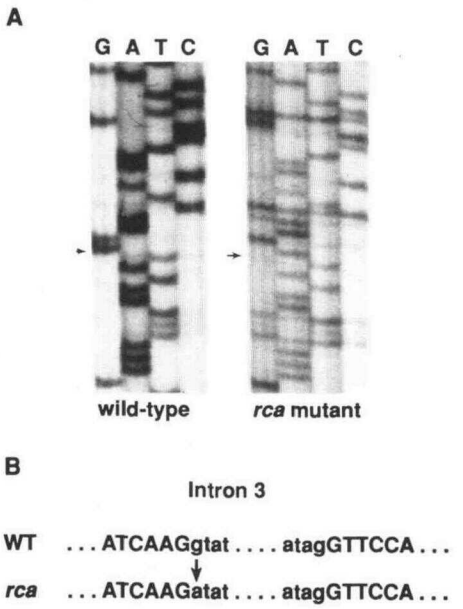

Figure 3. Sequence analysis of the 5′-splice junction of intron 3 in wild-type and *rca* genomic DNA. A, Nucleotide sequencing gels; B, nucleotide sequences. Arrows indicate the position of the G-to-A transition at the 5′ border in the *rca* mutant.

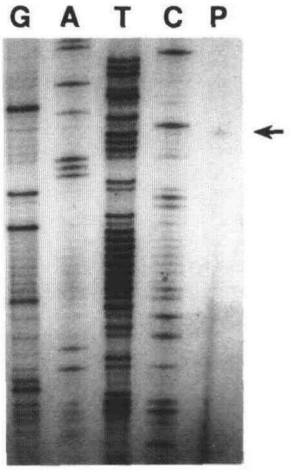

Figure 4. Primer extension analysis of the intron 3 region of *rca* Rubisco activase mRNA. The product of the primer extension reaction (P) was electrophoresed adjacent to dideoxy DNA-sequencing reactions (G, A, T, C) using the same ^{32}P-labeled primer. Extension was blocked at a putative lariat branch point (arrow).

stream from intron 3 to determine the 5′ end of the 1.8-kb transcript (Fig. 4). The primer extension reaction terminated at a U residue (nucleotide 2518 in Fig. 1), adjacent to a putative lariat formation site (TTGAT, nucleotides 2512–2516 in Fig. 1). No band was visible at the position corresponding to the 5′ end of intron 3. These results suggest that a lariat structure in the mRNA may have blocked reverse transcriptase from proceeding through the branch point.

DISCUSSION

The coding sequences of most eukaryotic genes contain introns that are removed by the process of mRNA splicing. In mammalian and yeast systems, intron splicing involves two transesterifications (Aebi et al., 1986). First, the 5′ end of the intron is cleaved and the 5′ phosphate is covalently linked to a 2′ hydroxyl within the intron, forming a lariat. Two intermediates are created, the linear upstream coding region (exon) and the lariat-intron attached to the downstream exon. In the second step, the 3′ end of the intron is cleaved and the two exons are joined. The products of the reaction are the spliced exons and the lariat intermediate of the intron.

Comparatively little is known about splicing of plant introns. However, more than 80% of plant protein-coding genes contain introns ranging in size from 70 nucleotides to more than 1000 nucleotides (Goodall and Filipowicz, 1989). Most characterized introns are flanked by a conserved GT at the 5′ end and conserved AG at the 3′ end (Goodall et al., 1991). Although lariat formation has not been demonstrated in plants, indirect evidence for branching was obtained by mutation of putative branch points in synthetic introns, which reduced splicing efficiency (Goodall and Filipowicz, 1989). Plant introns do not contain a strong conserved sequence for branch point selection, but a weak consensus similar to vertebrate branch points is located 20 to 40 nucleotides upstream from the 3′-splice site (Brown, 1986).

The sequences of genomic and cDNA clones isolated from *rca* mutant *Arabidopsis* revealed that an alteration in the 5′-splice junction of intron 3 was responsible for the rca phenotype. The cDNA clones obtained by library screening contained normal 3′ sequences. However, no full-length clones were obtained, and each of the cDNAs terminated within intron 3 sequences near the 5′ end. Sequence analysis of two independent cDNA clones generated by reverse transcription and PCR amplification showed that these also contained intron 3 and that a mutation had occurred at the 5′-splice site that converted the consensus GT dinucleotide to AT. Comparison of wild-type and *rca* genomic clones confirmed that a G-to-A transition had occurred at this position.

Northern analysis showed that three aberrant messages are produced as a result of the *rca* mutation. The presence of two bands corresponding to exons 1 through 3 and exons 4 through 7 indicate that the 5′ end of the intron was cleaved, but further processing was disrupted. A third band approximately 100 bp larger than wild type was found to be the full-length cDNA containing intron 3, indicating that efficiency of cleavage at the 5′ end was reduced. There was no evidence that cryptic splice sites were utilized. No additional bands were observed in northern analysis, and only the single product from the unspliced precursor was generated by PCR amplification across this region of the mRNA.

Possible structures of the three aberrant mRNAs found in the *rca* mutant, based on the analyses described above, are given in Figure 5. These structures are consistent with reports of similar mutations in yeast. In normal intron splicing, pre-mRNA is first cleaved at the 5′ junction at a conserved GT, producing a linear upstream segment and a downstream lariat intermediate in which the 5′ end of the intron is covalently linked to an A residue near the 3′-splice site (Aebi et al., 1986). The GT consensus is not required for cleavage at the 5′ junction or lariat formation but is necessary for subsequent cleavage at the 3′ junction in yeast and mammalian systems (Aebi et al., 1986; Vijayraghavan et al., 1986). Mutations in the conserved 5′ GT dinucleotide in yeast introns result in accumulation of splicing intermediates including precursor mRNA, the linear upstream moiety, and the lariat intermediate attached to the downstream coding region. In contrast, mammalian systems are able to activate cryptic splice sites in the intron and adjacent exon or to utilize the 5′-GT dinucleotide from the preceding intron. In this case, aberrant mRNAs are produced that are larger or smaller than the wild-type message, depending on where the alternative splice sites are located.

Northern analysis and the cDNA sequence analysis dem-

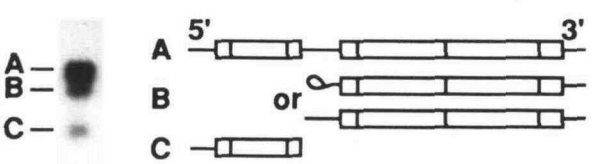

Figure 5. Proposed structures of Rubisco activase mRNA transcripts in the *rca* mutant. Wide areas indicate exons, vertical lines indicate positions of the introns, and horizontal lines indicate untranslated mRNA regions.

onstrated that the *rca* mutation results in aborted splicing after cleavage of the intron at the 5′ end, but it was unclear whether the 3′ intermediate moiety (1.8-kb mRNA) contains the entire intron 3 or was present in lariat form. Primer extension into the intron stopped at a U residue at a putative lariat branch point, suggesting that the message was in a lariat form. However, several cDNA clones extended to the 5′ end of the intron and terminated at a poly(T) region, whereas no cDNAs were isolated that ended near the branch point. Because full-length clones were also not recovered from the library, reverse transcription from the 2.0-kb mRNA may simply have terminated at the poly(T) tract during first-strand synthesis of the cDNA. Attempts to perform primer extension reactions from sites internal to the intron also terminated in the poly(T) region, which supports this interpretation. Thus, it is certain that the 1.8-kb message contains exons 4 to 7 and most or all of intron 3, but it is not known whether the intron is linear or in the form of a lariat. A comparison of the mRNAs preceding and following treatment with a debranching enzyme could clarify this point (Reed and Maniatis, 1985).

The *rca* mutant was obtained by treatment of *Arabidopsis* seeds (Somerville et al., 1982). The primary mode of action is alkylation of guanine to methylguanine, an analog of adenine (Kreig, 1963). During replication, methylguanine is paired with thymine, and the original guanine is subsequently replaced by adenine in the next round of replication. This is consistent with the findings in the *rca* mutant.

The data presented here demonstrate that the *Arabidopsis* Rubisco activase phenotype is caused by a mutation in intron 3 of the *rca* gene. A conversion of the consensus GT dinucleotide to AT at the 5′-splice junction results in inefficient splicing of the mRNA. No accurate wild-type transcripts are produced and, therefore, no Rubisco activase polypeptides result. The *rca* mutant has a stably inherited mutation that leads to the accumulation of mRNA splicing intermediates. Future analysis of these intermediates may provide valuable information about RNA-processing mechanisms in plants. Additionally, because the rca strain is a Rubisco activase null mutant, it provides an excellent system in which to systematically and directly test mutant enzymes prepared by directed mutagenesis for their effect in vivo after transformation and regeneration and for determining the importance of Rubisco activase in regulating photosynthesis.

Received November 24, 1992; accepted February 1, 1993.
Copyright Clearance Center: 0032-0889/93/102/0227/06.
The GenBank accession number for the sequence reported in this article is M86720.

LITERATURE CITED

Aebi M, Horning R, Reiser J, Weissmann C (1986) Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. Cell **47:** 555–565

Brown JWS (1986) A catalogue of splice junction and putative branch point sequences from plant introns. Nucleic Acids Res **14:** 9549–9559

Goodall GJ, Filipowicz W (1989) The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. Cell **58:** 473–483

Goodall GJ, Kiss T, Filipowicz W (1991) Nuclear RNA splicing and small nuclear RNAs and their genes in higher plants. Oxford Surv Plant Mol Cell Biol **7:** 255–296

Kawasaki ES (1990) Amplification of RNA. *In* MA Innis, DH Gelfand, JJ Sninsky, TJ White, eds, PCR Protocols: A Guide to Methods and Applications. Academic Press, San Diego, CA, pp 21–27

Krieg DR (1963) Ethylmethanesulfonate-induced reversion of bacteriophage T4 rII mutants. Genetics **48:** 561–580

Lorimer GH, Miziorko HM (1980) Carbamate formation on the ω-amino group of a lysyl residue as the basis for the activation of ribulose-bisphosphate carboxylase by CO_2 and Mg^{2+}. Biochemistry **19:** 5321–5328

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular Cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Ogren WL (1984) Photorespiration: pathways, regulation, and modification. Annu Rev Plant Physiol **35:** 415–422

Portis AR Jr, Salvucci ME, Ogren WL (1987) Activation of ribulose bisphosphate carboxylase/oxygenase at physiological CO_2 and ribulose bisphosphate concentrations by rubisco activase. Plant Physiol **82:** 967–991

Reed R, Maniatis T (1985) Intron sequences involved in lariat formation during pre-mRNA splicing. Cell **41:** 95–105

Rundle SJ, Zielinski RE (1991) Organization and expression of two tandemly oriented genes encoding ribulosebisphosphate carboxylase/oxygenase activase in barley. J Biol Chem **266:** 4677–4685

Salvucci ME, Portis AR Jr, Ogren WL (1985) A soluble chloroplast protein catalyzes ribulose bisphosphate carboxylase/oxygenase activation *in vivo*. Photosynth Res **7:** 193–201

Salvucci ME, Werneke JM, Ogren WL, Portis AR Jr (1987) Purification and species distribution of rubisco activase. Plant Physiol **84:** 930–936

Somerville CR, Portis AR Jr, Ogren WL (1982) A mutant of *Arabidopsis* which lacks activation of RuBP carboxylase *in vivo*. Plant Physiol **70:** 381–387

Verwoerd TC, Dekker BMM, Hoekama A (1989) A small scale procedure for the rapid isolation of plant RNAs. Nucleic Acids Res **17:** 2362

Vijayraghavan U, Parker R, Tamm J, Iimura Y, Rossi J, Abelson J, Guthrie C (1986) Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway, particularly assembly of the spliceosome. EMBO J **5:** 1683–1695

Werneke JM (1989) Structure and function of *Spinacea oleracea* L. and *Arabidopsis thaliana* cDNAs encoding ribulosebisphosphate carboxylase/oxygenase. PhD Dissertation, University of Illinois, Urbana, IL

Werneke JM, Chatfield JM, Ogren WL (1989) Alternative mRNA splicing generates the two rubisco activase polypeptides in spinach and *Arabidopsis*. Plant Cell **1:** 815–825

Werneke JM, Ogren WL (1989) Structure of an *Arabidopsis thaliana* cDNA encoding rubisco activase. Nucleic Acids Res **17:** 2871

Werneke JM, Zielinski RE, Ogren WL (1988) Structure and expression of spinach leaf cDNA encoding ribulosebisphosphate carboxylase/oxygenase activase. Proc Natl Acad Sci USA **85:** 787–791